Virtual Camera Perspectives within a 3-D Interface for Robotic Search and Rescue

David Bruemmer, Douglas Few, Heather Hunting, Miles Walton
Human, Robotic, and Remote Systems Department
Idaho National Laboratory
Idaho Falls, ID, U.S.A.
{david.bruemmer, douglas.few, heather.hunting, miles.walton}@inl.gov

Curtis Nielsen
Machine Intelligence, Learning, and Decisions Lab
Brigham Young University
Provo, Utah, USA
curtisn@cs.byu.edu

Abstract *– This paper discusses the use of four different camera perspectives within a real-time virtual 3-D interface as it is used to accomplish a remote robotic exploration and mapping task. The 3-D interface is used as the basis for a Cognitive Collaborative Workspace (CCW) that supports shared understanding of the task and environment. Multiple humans and robots can add iconographic entities such as waypoints, areas of interest, start and end locations, humans, doors, and landmines into the CCW. These tools can be used as a basis for tasking and monitoring and also communicate intentions and percepts. The experiment discussed in this paper evaluates how different perspectives within the 3-D display affect the performance of human-robot teams on a map-building and exploration task. Results show that perspective does play an important role in providing situation awareness. Specifically, task efficiency in terms of time to complete task and joystick usage is significantly diminished when the endocentric (i.e. 1st person) perspective is used.*

I. INTRODUCTION

In many hazardous and critical environments, it is not possible to maintain continuous, high-bandwidth communication. For instance, attempts to use mobile robots at the world trade center were severely limited by the fact that communications could not reach reliably into crevices and crawl spaces [1]. Even under normal conditions, the utility of teleoperated robot control strategies is limited by an inability to transmit video far inside of concrete structures. To move beyond the limitations of teleoperation, this research reports on the use of an abstracted representation that replaces video and the accompanying need for high-bandwidth communication. Previous experiments have shown that this virtual 3-D interface can reduce operator error and workload and effectively remove the need for video. [2,3] Instead of video, a real-time map used as the basis for an abstracted 3-D video-game representation. By sending only new range abstractions, this interface utilizes up to 5000 times less bandwidth than a teleoperated control strategy. The resulting cognitive collaborative workspace (CCW) fuses video into a virtual 3-D representation of the world that can be used to navigate unknown, dynamic environments and build a shared representation that promotes situation awareness and intuitive tasking. Although the virtual 3D display offers benefits in terms of reduced bandwidth, operator workload and error, it is not clear how perspective influences these effects.

An increasing number of researchers from the fields of human factors, cognitive science, and robotics are working to develop new HRI methods for remote operation of mobile vehicles (see [4] for an overview). Casper and Murphy present a post-hoc analysis of the rescue efforts at the World Trade Center in September 2001 where robots were used for the first time to assist in real, un-staged search and rescue operations [1]. Burke et al. present a field study on human-robot interaction in an urban search and rescue training task [5]. Yanco et al. [6] present an analysis of the 2002 American Association for Artificial Intelligence (AAAI) Robot Rescue Competition where robot systems were used to compete in a mock search and rescue operation. In each study, the authors noted that it was difficult for operators to navigate due to an inability to understand the robot's position and/or perspective within the remote environment. Unlike video, which offers only a 1st-person, local environment

perspective, the 3-D interface can change perspective to support different levels of robot autonomy and different elements of a task. For instance, navigation may require a different perspective than a visual inspection task where the robot remains stationary while a camera is panned and tilted to survey the local environment. The purpose of this study is to investigate the role of perspective in terms of operator error, workload and overall task efficiency.

II. SYSTEM DESIGN

II.A. Robot Implementation

The control architecture discussed in this article is the product of a spiral development cycle where behaviors have been evaluated in the hands of users, modified, and tested again. The INL has developed a behavior architecture that can port to a variety of robot geometries and sensor suites and which is being used as a standard by several HRI research teams throughout the community. Experiments discussed in this paper utilized the iRobot "ATRV mini" shown in Figure 1. The behavior architecture utilizes a variety of sensor information including inertial sensors, compass, wheel encoders, laser, computer vision, tilt sensors, and a full ring of ultrasonic sensors.

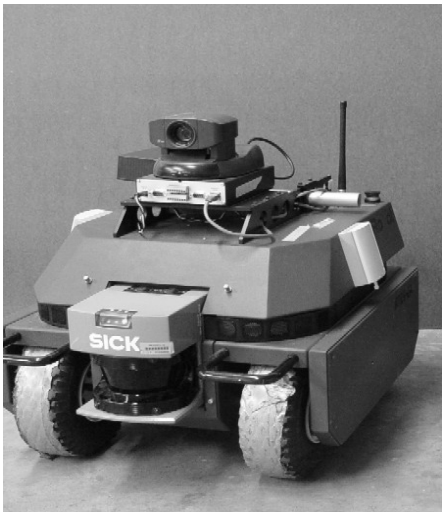

Fig. 1: The robot used for this experiment.

Using a technique described in [7], a *guarded motion* behavior permits the robot to take initiative to avoid collisions. In response to laser and sonar range sensing of nearby obstacles, the robot scales down its speed using an event horizon calculation, which measures the maximum speed the robot can safely travel in order to come to a stop approximately two inches from the obstacle. By scaling down the speed by many small increments, it is possible to insure that regardless of the commanded translational or rotational velocity, guarded motion will stop the robot at the same distance from an obstacle. This approach provides predictability and ensures minimal interference with the operator's control of the vehicle. If the robot is being driven near an obstacle rather than directly towards it, *guarded motion* will not stop the robot, but may slow its speed according to the event horizon calculation.

II.B Virtual 3D Display

The goal of the 3-D display is to provide a workspace for collaborative understanding between the human and robot. The virtual 3-D component has been developed by melding technologies from the INL [2], Brigham Young University (BYU) [2], and Stanford Research Institute (SRI) International [8,9]. The 3D virtual display is not based on true 3-D range sensing, but rather by extruding a 2D map to provide the user with a malleable perspective. To build the map, the INL control system uses a technique developed at SRI called Consistent Pose Estimation (CPE) that allows for efficient incorporation of new laser scan information into a growing map. CPE also addresses the problem of loop closure: how to register new laser information when the robot returns to a previously explored area.

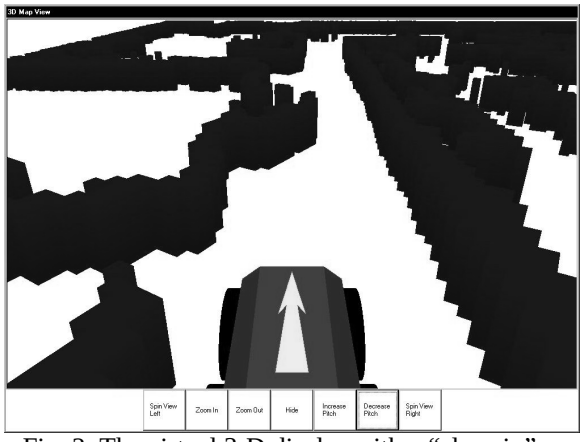

Fig. 2: The virtual 3-D display with a "close-in" perspective

The map produces the basis for the 3-D representation that includes obstacles and other semantic entities that are of significance to the operator such as start location, labels and waypoints. These items can be inserted by the robot to indicate percepts or intentions; likewise, the operator may insert entities from a drop down menu. The operator may also insert translucent still images, excerpted from the robot video, which are overlaid onto the corresponding area of the 3D map display, providing a means to fuse real and virtual elements. By changing the zoom and pitch of the interface field of view, it is possible to move from an egocentric perspective (i.e. looking out from the robot), to a fully exocentric view where the entire environment can be seen

at once. The present study investigates this spectrum in terms of overall system performance, navigational error (instances of robot initiative) and workload (joystick usage).

III EXPERIMENT

III.A Participants

The present study included 216 participants drawn at random from attendees of the Idaho National Laboratory's (INL'S) annual community exposition. The participants consisted of 61 females and 155 males, ranging in age from 3 to 70 years old, with a mean age of 12.

III.B Robot Description

The robot used in the present study was a wheeled ATRVmini manufactured by iRobot, which measures approximately 67 cm long x 54 cm wide including the wheelspan. The sensory information is used by the robot itself to take initiative during the task and is also available to the robot operator in the form of meaningful abstractions. The robot can provide a video feed to the operator from a forward mounted camera. In the standard configuration, the video signal and the sensory data are fed to the control station via a wireless link and superimposed onto the 3-D display so that the live video feed corresponds to the map abstractions. However, for this experiment the video was not shown to the operators.

III.C Interface Description

In this experiment, control of the robot was achieved by manipulating the joystick. The participants were instructed to turn or push the joystick in the direction they wished the robot to move. Participants were informed that if the robot took initiative to prevent a collision, the joystick would vibrate to inform them that motion was blocked in that direction. The interface logged the total number of joystick commands given which will be referred to as total joystick bandwidth. The robot autonomy was configured to allow the participant to direct the robot with the joystick, but the robot was enabled to take initiative to slow or stop itself as necessary to avoid collisions. The measure of how often the robot had to come to a full stop to prevent a collision was recorded by the interface and will be referred to as robot initiative. Note that robot initiative is also a metric of human navigational error. The interface as displayed to the participants, consisted of a three-dimensional map of the maze built as the robot explored the environment. Four different perspectives of this map were available. Throughout the experiment, each volunteer used one of these four perspectives. *1st person perspective* places the camera inside the robot, so the view is what it would be if the participant was sitting in the robot. It is analogous to the perspective provided by a normal teleoperation interface where the user sees the video from the perspective of the robot's camera. *Close perspective* is zoomed out slightly and uses a virtual camera position behind the robot such that the front half of the robot is also visible at the bottom of the screen. Fig. 2 illustrates the vantage point seen using the close perspective. *Elevated perspective* zooms the map display out and places the camera behind and above the robot. See Fig. 3 for an example of elevated perspective. Note that Fig. 3 also includes an autonomously generated waypoint path plan as well as a snap shot left by the robot. *Far perspective* zooms out further by placing the virtual camera position directly above the robot. It is far enough above the robot to allow the entire map to be visible on the screen. This is often referred to as a "Bird's Eye View."

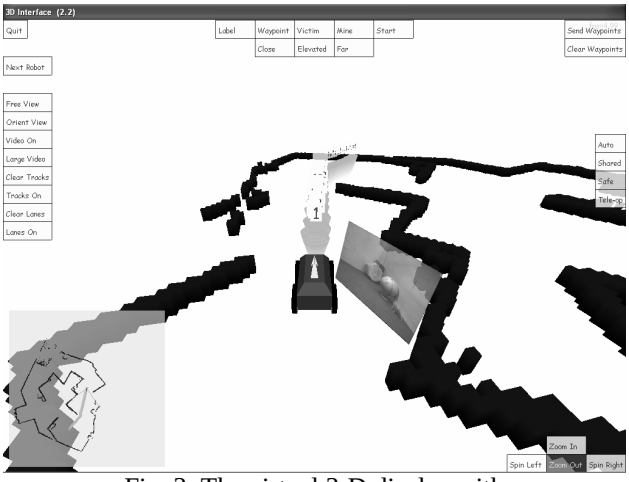

Fig. 3: The virtual 3-D display with an "Elevated" perspective

III.D. Environment

On the first floor of the Museum of Idaho, a maze environment was constructed using cubicle wall dividers. On the second floor of the museum, a control station was constructed that consisted of a laptop and monitor to display the interface and a joystick with which to control the robot. The participants could see the interface, but did not have any ability to see the actual robot or the maze itself.

III.E.Procedure

Each participant was instructed on the use of the joystick for controlling the robot. They were then requested to build a complete map of the maze as quickly as possible without running the robot into obstacles. Participants were also informed that the robot would prevent collisions, but that they should drive the robot in order to prevent such instances. Each participant used one

of the four perspectives, which were assigned to volunteers in successive, cyclical order. Information including the time required to complete the task, the initiative exercised by the robot, and the total joystick bandwidth used to guide the robot was measured and recorded automatically and stored in a data file on the interface computer. Also, information on age, gender and a self-assessment of video game skill (on a scale of 1 to 10) was also recorded for each participant. .

III.F Results

The effects of participant age, gender, self-rated video game skills, and perspective were compared against the time it took to build a complete map, the initiative exercised by the robot, and the total joystick bandwidth used to communicate with the robot.

III.F.1 Time

The participants were grouped by age in three-year intervals. There was a significant difference in the time required to complete the map due to age, $F(11, 115) = 2.715$, $p = 0.004$. Note that the study did not feature a balanced sample across age groups. More young participants volunteered than did older adult participants. The results indicate that there was a tendency for teenagers to complete the map more quickly, however, due to the lack of a balanced sample across all age groups, it is not possible to draw any definitive conclusions about the interaction of time and age.

There was no significant difference in time due to gender, $F(1, 116) = 3.16$, $p = 0.078$. Females took on average 113.98 seconds to complete the map, while males took on average 101.86 seconds. There was no significant difference in time due to video game skills, $F(4, 116) = 2.219$, $p = 0.071$. There was a significant difference in time due to perspective, $F(3, 116) = 13.632$, $p < 0.001$. Participants using 1st person perspective took on average 133.38 seconds, while close, elevated, and far had averages of 95.29, 96.43, and 96.76 seconds, respectively.

III.F.2 Initiative

The same three-year age groups were used. There were no significant differences in initiative exercised by the robot due to age, $F(10, 87) = 1.141$, $p = 0.342$. There were no statistical differences in initiative due to gender, $F(1, 87) = 0.250$, $p = 0.619$. There were no statistical differences in initiative due to video skills, $F(4, 87) = 1.202$, $p = 0.316$. There were no statistical differences in initiative due to perspective, $F(3, 87) = 0.383$, $p = 0.766$. There were also no significant interactions among the variables.

III.F.3 Total Joystick Bandwidth

Again, the same three-year age groups were used. There was a significant difference in total joystick bandwidth due to age, $F(10, 87) = 1.835$, $p = 0.066$. The data indicates a tendency for teenagers to use less the joystick; however, lack of a balanced sample set prevents definitive assertions regarding age. There was a significant difference due to mode, $F(3, 87) = 16.442$, $p < 0.001$. The participants using the 1st person perspective used significantly more joystick bandwidth to control the robot than the other three perspectives. The mean for 1st person was 1344.51, compared to 763.75, 724.07, and 693.09 for close, elevated, and far, respectively.

There was a significant difference due to gender, $F(1, 87) = 6.674$, $p = 0.011$. Males used an average bandwidth of 840.80, while females used an average of 963.88. Thus males appeared to be more efficient in controlling the robot. There was a significant interaction involving Gender*Video Skills, $F(4, 87) = 3.155$, $p = 0.018$. For the most part, the higher that males rated their video skills, the more bandwidth they used to control the robot; females, on the other hand, used less bandwidth when they rated themselves higher.

IV. CONCLUSIONS

This experiment indicates that perspective is important when remotely operating a robot. Previous experiments [1,2] had shown that a virtual 3-D display could remove the need for continuous video, increase overall task efficiency and reduce operator error and workload. However, it was unclear what role perspective had in bringing about these benefits. It was possible that the benefits due to the 3-D perspective were largely due to the simplification brought through the abstraction process. However, it was also possible that the main benefit of the 3-D display was that it provided a perspective that was more useful for navigation and exploration missions than the traditional video display used in teleoperation. This study was intended to serve as a preliminary investigation into this question. Would the benefits of the 3-D display be seen across all perspectives or would the issue of perspective prove to be a critical factor in determining the utility of the 3-D display?

The results presented here indicate that the 1st person perspective within the 3-D display, which uses a similar perspective as the presentation of video within a traditional interface, is inferior to the exocentric perspectives that show the robot and how it fits into the world. Although perspective is a critical factor in terms of time and joystick usage, it does not, at least for this study, seem to play a critical role in terms of operator navigational error (i.e. instances which necessitated robot

initiative). It is perhaps not surprising that perspective plays an important role; but what is surprising is that once the perspective moves from the 1st person to include the robot, there seems to be little difference between the various exocentric perspectives used. Close, Elevated and Far all seemed to be very similar in terms of time, joystick usage, and robot initiative. Additional studies will be necessary to further understand the benefits and limitations associated with different perspectives. Most likely, there will not be one optimal perspective. Rather perspective should change based on the task element (e.g. navigation, search, patrol), the level of robot autonomy (e.g. direct human control, shared control, autonomous tasking) and the number of robots employed.

REFERENCES

1. J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 33(3):367–385, 2003.
2. Bruemmer, D.J. Few, D.A. Boring, R.L. Marble, J.L. Walton, M.C. Nielsen, C.W. "Shared understanding for collaborative control." *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 35(4): 494-504, July 2005.
3. D. J. Bruemmer, D. A. Few, R. Boring, M. Walton, J. L. Marble, C. Nielsen, J. Garner. "'Turn Off the Television!' Robotic Exploration Experiments with a 3D Abstracted Map Interface," In *Proceedings of the 38th Hawaii International Conference on the System Sciences*, Waikoloa Village, Hawaii, January 2005.
4. J. L. Burke, R R. Murphy, M. D. Coovert, and D. L. Riddle. "Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise," *Human-Computer Interaction*, 19:85–116, 2004.
5. J. L. Burke, R. R. Murphy, Erika Rogers, V. J. Lumelsky, and J. Scholtz. "Final report for the DARPA/NSF interdisciplinary study on human-robot interaction," *IEEE Transactions on Systems, Man, and Cybernetics*, Part C, 34(2):103–112, May 2004.
6. H. A. Yanco, J. L. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Journal of Human-Computer Interaction*, Vol. 19, pp. 117-149, 2004.
7. E.B. Pacis, H.R. Everett, N. Farrington, D. J. Bruemmer, "Enhancing Functionality and Autonomy in Man-Portable Robots," In *proceedings of the SPIE Defense and Security Symposium 2004.* 13 -15 April, 2004
8. K. Konolige. "Large-scale map-making," *In Proceedings of the AAAI,* San Jose, CA, 2004.
9. J.S Gutman and K. Konolige. "Incremental Mapping of Large Cyclic Environments," In *proceedings of the CIRCA 99*, Monterey, California, 1999.